Strong Shift Equivalence and Shear Adjacency of Nonnegative Square Integer Matrices

Kirby A. Baker
Department of Mathematics
University of California
Los Angeles, California 90024

Submitted by Hans Schneider

ABSTRACT

For two square matrices A, B of possibly different sizes with nonnegative integer entries, write $A \approx_1 B$ if $A = RS$ and $B = SR$ for some two nonnegative integer matrices R, S. The transitive closure of this relation is called *strong shift equivalence* and is important in symbolic dynamics, where it has been shown by R. F. Williams to characterize the isomorphism of two topological Markov chains with transition matrices A and B. One invariant is the characteristic polynomial up to factors of λ. However, no procedure for deciding strong shift equivalence is known, even for 2×2 matrices A, B. In fact, for $n \times n$ matrices with $n > 2$, no nontrivial sufficient condition is known. This paper presents such a sufficient condition: that A and B are in the same component of a directed graph whose vertices are all $n \times n$ nonnegative integer matrices sharing a fixed characteristic polynomial and whose edges correspond to certain elementary similarities. For $n > 2$ this result gives confirmation of strong shift equivalences that previously could not be verified; for $n = 2$, previous results are strengthened and the structure of the directed graph is determined.

1. INTRODUCTION

In symbolic dynamics, the topological Markov chain (subshift of finite type) derived from a finite directed graph $G = (V, E)$ is the pair (S_G, σ), where S_G is the space of two-way-infinite walks on G, as $\mathbf{Z}$-indexed sequences of edges (i.e., as elements of $E^{\mathbf{Z}}$), and where $\sigma : S_G \to S_G$ is the left shift as a self-homeomorphism of S_G with respect to the topology inherited from the product topology on $E^{\mathbf{Z}}$ [24, 2]. Topological Markov chains have important applications to coding theory [1, 12], ergodic theory [9], and the study of

52 Vanderbilt Ave., New York, NY 10017 0024-3795/87/$3.50

diffeomorphisms [4, 5]. Each such space is determined by the transition matrix (adjacency matrix) of the directed graph, a square nonnegative integer matrix A, where a_{ij} is the number of edges from vertex i to vertex j.

A basic question is how to determine whether two given topological Markov chains are topologically conjugate (isomorphic, in the sense of being homeomorphic by a map commuting with left shifts). A fundamental theorem of R. F. Williams states a criterion: if and only if their transition matrices are strongly shift equivalent in the following sense:

DEFINITION 1.1. For two square nonnegative integer matrices A, B, write $A \approx_1 B$ if there exist nonnegative integer matrices R, S with $A = RS$, $B = SR$. A and B are *strongly shift equivalent*, here denoted $A \approx B$, if they are connected by a finite sequence $A = A_0 \approx_1 A_1 \approx_1 \cdots \approx_1 A_k = B$. Here A, B, and the intermediate matrices can be different square sizes, and k can depend on A and B.

No algorithm for determining strong shift equivalence is known, even for 2×2 matrices. For example, it is not even known whether

$$\begin{bmatrix} 2 & 3 \\ 3 & 0 \end{bmatrix} \approx \begin{bmatrix} 1 & 2 \\ 5 & 1 \end{bmatrix}.$$

For this reason, conditions that are either necessary or sufficient are of importance.

An obvious necessary condition for $A \approx B$ is that A and B have the same characteristic polynomial up to factors of λ (or equivalently, have the same zeta function [6]); a refinement of this observation is that A and B must have the same Jordan form "away from 0" [7]. A sharper necessary condition is that A and B are "shift equivalent": there exist nonnegative integer matrices R, S and an integer $k \geqslant 1$ such that $A^k = RS$, $B^k = SR$ and such that $SA = BS$, $AR = RB$ [24]. Shift equivalence is computable [13, 14]. "Williams's conjecture" is that shift equivalence and strong shift equivalence coincide.

A sufficient condition for strong shift equivalence of 2×2 integer matrices A, B of nonnegative determinant was given in [2]: It suffices for A and B to be similar over the integers, i.e., $P^{-1}AP = B$ for an integer matrix P of determinant ± 1. (Previously, Cuntz and Krieger [8] had proved the cases for which $\det A = \pm 1$.) This theorem solved many instances of conjectured strong shift equivalence and generated new examples. The main tool used was similarity via unit shears (as defined in Definition 2.1 below).

The purpose of this paper is twofold: (1) to extend the theory of similarity by unit shears to the $n \times n$ case, and (2) simultaneously to determine the structure of the 2×2 case.

For (1), it is shown in Sections 2, 3 that similarity via unit shears for $n \times n$ matrices with $n > 2$ is still a sufficient condition for strong shift equivalence, with only a mild nontriviality condition (Definition 2.3, Lemma 3.1). An explicit algorithm is given, in which the intermediate matrices are $(n+1)\times(n+1)$. It is natural to use similarity via unit shears to yield a digraph (directed graph) structure on the space of nonnegative square integer matrices with a fixed characteristic polynomial—a "shear digraph." Each strong shift equivalence class is then a union of components of this digraph (Theorem 3.2). This fact can be used to demonstrate the strong shift equivalence of many examples of pairs of matrices for which previously it could only be conjectured, and it provides a framework for future study of Williams's conjecture.

For (2), it is shown in Sections 4–5 that the shear digraph is especially tractable in the 2×2 case, and the possible digraphs of this kind are determined (Corollary 4.2, Theorem 4.6, Theorem 5.9). A by-product is the extension of the result of [2], quoted above, to nonnegative 2×2 integer matrices the sum of whose determinant and trace is nonnegative (Theorem 5.6).

General references for this theory are [24, 10, 16, 17, 22, 19]. General references for the relevant properties of integer matrices are [20, 23] and for nonnegative matrices [11]. A reference for graph theory is [3].

2. UNIT SHEARS AND THE SHEAR DIGRAPH

DEFINITION 2.1. A *unit shear* (or *unit transvection matrix*) is a square matrix whose entries are those of the identity matrix except for one extra off-diagonal 1. For $i \neq j$, let $U_n[i, j]$ denote the $n \times n$ unit shear in which the extra 1 is in the (i, j) position.

OBSERVATION 2.2 [21]. *$U_n[i, j]$ and $U_n[k, l]$ commute except when $j = k$ or $i = l$. If $j = k$ and i, j, l are distinct, then the commutator of $U_n[i, j]$ and $U_n[j, l]$ is $U_n[i, l]$, which then commutes with both. (Here, as in [21], the appropriate version of the commutator of S and T is $[S, T] = STS^{-1}T^{-1}$. Actually, $U_n[i, j]$ and $U_n[j, l]$ generate a free class-2 nilpotent group; i.e., commutators are central.) The case $i = l$ is of course similar, but with commutator $U_n[k, j]^{-1}$.*

DEFINITION 2.3. For $n \times n$ nonnegative integer matrices A and B, write $A \to B$ if there is a unit shear $U = U_n[i, j]$ such that (i) $U^{-1}AU = B$, (ii)

$a_{ij}, b_{ij} > 0$, or equivalently, $A, B \geqslant U - I$. A and B are *shear adjacent* if $A \to B$ or $B \to A$.

DEFINITION 2.4. The *shear digraph* $\mathrm{SDG}_{p(\lambda)}$ of a monic polynomial $p(\lambda) \in \mathbf{Z}[\lambda]$ is the directed graph, possibly infinite, whose vertices consist of all square nonnegative integer matrices with characteristic polynomial $p(\lambda)$ and whose edges are determined by the $\to$ relation. The shear digraph SDG_A of a square nonnegative integer matrix A is the shear digraph of its characteristic polynomial. (Informally, let us use this notation both for a shear digraph and its set of vertices.)

A shear digraph can have loops, but no multiple loops or multiple edges. Of course, a shear digraph can be empty. Example 3.5 shows one component of a shear digraph of 3×3 matrices; Example 4.8 shows a shear digraph of 2×2 matrices that is a disjoint union of cycles.

Recall that a nonnegative square matrix is *irreducible* if no conjugation by a permutation results in a nontrivially block-triangular matrix [11, p. 50]. For an integer matrix, an equivalent statement is that the corresponding digraph is strongly connected, in the sense that there is a directed path from each vertex to any other.

PROPOSITION 2.5 (Cf. [7]). *A shear digraph has only finitely many irreducible matrices among its vertices.*

In particular, if $p(\lambda)$ is monic and irreducible in $\mathbf{Z}[\lambda]$, then its shear digraph has only irreducible matrices as vertices and so is finite.

Proof of Proposition 2.5 (as observed by M. Boyle). If A is $n \times n$ with characteristic polynomial $p(\lambda)$, then the eigenvalues are determined and hence $\mathrm{trace}(A^k)$ is determined for $k = 1, 2, \ldots$. If A is irreducible, then every entry of A contributes to $\mathrm{trace}(A^k)$ for some $k \leqslant n$. If in addition A is nonnegative with integer entries, then $\max_{k \leqslant n} \mathrm{trace}(A^k)$ is a bound on the entries of A. ■

3. SHEAR ADJACENCY AND STRONG SHIFT EQUIVALENCE

LEMMA 3.1. *Let A and B be $n \times n$ nonnegative integer matrices. If $A \to B$, then $A \approx B$.*

The proof is given below after Lemma 3.3.

THEOREM 3.2. *In a shear digraph, any two matrices in the same component are strongly shift equivalent.*

(Two matrices are said to be in the same strong component if they are connected by a directed path in a graph; they are in the same (weak) component if they are in the same component of the corresponding undirected graph, in which edges are the same but edge directions are ignored.)

LEMMA 3.3. *For an $n \times n$ matrix M, let $\overline{M}$ be the $(n+1)\times(n+1)$ matrix obtained by bordering M with an extra row and column of zeros, except for a 1 in the $(1, n+1)$ position. Then $\overline{M} \approx_1 M$.*

Proof.

$$M = [M \quad \mathbf{e}]\begin{bmatrix} I \\ 0 \end{bmatrix}; \ \overline{M} = \begin{bmatrix} I \\ 0 \end{bmatrix}[M \quad \mathbf{e}],$$

where $\mathbf{e}$ is the column vector with entries $1, 0, \dots, 0$. ∎

Proof of Lemma 3.1. By Lemma 3.3 we need show only that $\overline{A} \approx \overline{B}$. We accomplish this by an explicit sequence of one-step strong shift equivalences via similarities, to obtain a strong shift equivalence of this form:

$$A \approx_1 \overline{A} \approx_1 M_{(0)} \approx_1 M_{(1)} \approx_1 \cdots \approx_1 M_{(m)} \approx_1 E \approx_1 \overline{B} \approx_1 B, \tag{3.4}$$

where $m = b_{11}$.

The matrices involved are defined as follows: We may assume that $H = U_n[1,2]$. Let $P = U_{n+1}[1,2]$ (a bordering of H), $Q = U_{n+1}[n+1,2]$, $L = \prod_{i \neq 2,\ i \leqslant n} U_{n+1}[n+1, i]^{b_{1i}}$. Thus L has the entries of the identity matrix except that its last row is $b_{11}, 0, b_{13}, b_{14}, \dots, b_{1n}, 1$. Now for $i = 0, 1, \dots, m$ let $M_{(i)} = Q^{-i} L \overline{A} L^{-1} Q^i$, and let $E = L\overline{B}L^{-1}$.

From Observation 2.2 it follows that the commutator $LPL^{-1}P^{-1}$ equals $Q^{b_{11}} = Q^m$. From this it follows that $P^{-1}Q^{-m}L = LP^{-1}$ and hence that $P^{-1}M_{(m)}P = E$.

Thus, instead of conjugating $\overline{A}$ directly by P to obtain $\overline{B}$, the successive steps represent conjugation by L^{-1}, m times by Q, then by P, and finally by L. This detour via L involves a noncommutativity with P that is balanced by the fact that their commutator is Q^m.

The one-step strong shift equivalences in detail are these:

(1) $\overline{A} \approx_1 M_{(0)}$ via $\overline{A} = (\overline{A}L^{-1})L$, $M_{(0)} = L(\overline{A}L^{-1})$;

for $i = 1, \ldots, m$,

(2_i) $M_{(i-1)} \approx_1 M_{(i)}$ via $M_{(i-1)} = Q(M_{(i)}Q^{-1})$, $M_{(i)} = (M_{(i)}Q^{-1})Q$;
(3) $M_{(m)} \approx_1 E$ via $M_{(m)} = P(P^{-1}M_{(m)})$, $E = (P^{-1}M_{(m)})P$;
(4) $E \approx_1 \bar{B}$ via $E = L(\bar{B}L^{-1})$, $\bar{B} = (\bar{B}L^{-1})L$.

In practice, this sequence of one-step strong shift equivalences can be shortened, in a way heavily dependent upon the particular matrices involved. Of course, if $B = U^{-1}AU$ with $U^{-1}A \geqslant 0$, then $A \approx_1 B$, so that the whole construction is unnecessary. For the general case, the first two steps can always be coalesced into one, as can the last two. Usually some, but not all, of the steps involving the $M_{(i)}$ can be coalesced; the actual minimum number of steps depends on the particular A and B in question.

To prove that the steps (1), (2_i), (3), (4) do indeed constitute a strong shift equivalence, it is necessary to show (a) that the intermediate matrices $M_{(i)}$ and E are nonnegative, and (b) that the factors $\bar{A}L^{-1}$, $M_{(i)}Q^{-1}$, $P^{-1}M_{(m)}$, and $\bar{B}L^{-1}$ are all nonnegative. Although it is evident that (b) implies (a), we proceed to derive instances of (b) and hence (a) individually:

(1′) $\bar{A}L^{-1}$ differs from $\bar{A}$ only in that the first row has been replaced by $a_{21}, a_{12}, a_{23}, a_{24}, \ldots, a_{2n}, 1$. Here we use the fact that the entries b_{1j} in the last row of L can be rewritten as $a_{1j} - a_{2j}$, because of the relation $P^{-1}\bar{A}P = \bar{B}$. A consequence [from (1)] is that $M_{(0)} \geqslant 0$.

(4′) Similarly, $\bar{B}L^{-1}$ differs from $\bar{B}$ only in that the first row has been replaced by $0, b_{12}, 0, \ldots, 0, 1$. A consequence [from (4)] is that $E \geqslant 0$.

(3′) $P^{-1}M_{(m)} = EP^{-1} = L\bar{B}L^{-1}P^{-1}$ is nonnegative because it is the product of L and $\bar{B}L^{-1}P^{-1}$, which differs from $\bar{B}L^{-1}$ only in that its second column has been replaced by $b_{12}, a_{22}, a_{32}, \ldots, a_{n2}, 0$. A consequence [from (3) and (4′)] is that $M_{(m)} \geqslant 0$.

(2_i): For $M_{(i)}Q^{-1}$ let us use a less direct argument:

First, observe that the last column of each of E and the $M_{(i)}$ will consist of entries that are 0, except for the top entry (which is 1) and the bottom entry.

Second, observe that $M_{(i)}Q^{-1}$ for $i = 1, \ldots, m$, if regarded as a single matrix function of i, is linear in each entry. Indeed, $M_{(i)}Q^{-1} \equiv Q^{-i}L\bar{A}L^{-1}Q^{i-1}$, which can be regarded as the result of (a) multiplying $L\bar{A}L^{-1}$ on the right by Q^{i-1} and then (b) multiplying this result on the left by Q^{-i}. Here (a) has the effect of adding $i - 1$ times the last column to the second; because of the zeros in the last column, this affects only the $(1,2)$ and $(n+1, 2)$ entries, and in particular not the second row. Then (b), which has the effect of subtracting i times the second row from the last row, likewise affects the entries only linearly in i. By this observation of linearity, it suffices to check the nonnegativity only at the extreme values of i, namely, $i = 1$ and $i = m$. In fact, we can be even more extreme and check the cases $i = 0$ and $i = m$.

Now observe that because Q commutes with L and P, $M_{(0)}Q^{-1}$ and $M_{(m)}Q^{-1}$ can be regarded respectively as the new $M_{(0)}$ and $M_{(m)}$ obtained by replacing $\bar{A}$ with $\bar{A}Q^{-1}$ and $\bar{B}$ with $\bar{B}Q^{-1}$, or equivalently, by decrementing a_{12} and b_{12} by 1 in A and B. By the definition of shear adjacency these new A and B are still nonnegative. The argument in (1′) and (3′) for the nonnegativity of $M_{(0)}$ and $M_{(m)}$, repeated for the new A and B, therefore serves to establish the nonnegativity of the old $M_{(0)}Q^{-1}$ and $M_{(m)}Q^{-1}$, as desired. Thus (2_i) holds for $i = 1, \ldots, m$.

This completes the verification that (3.4) is indeed a strong shift equivalence of A and B. ■

It should be noted that the proof just concluded required careful handling of the elements not in rows 1, 2 and columns 1, 2, an aspect not present in [2].

EXAMPLE 3.5. Figure 1 shows one component of a shear digraph. The vertices are all strongly shift equivalent, a fact that would not be evident without Theorem 3.2.

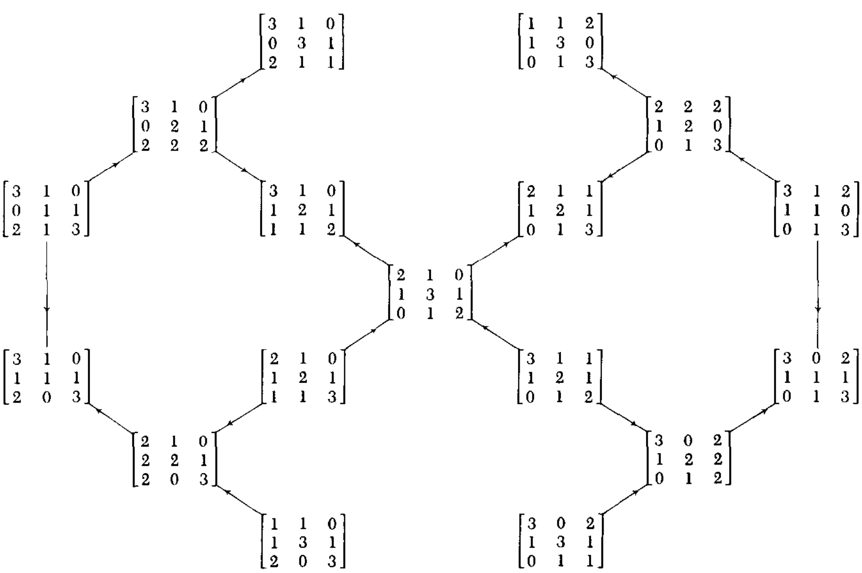

FIG. 1. One component of a shear digraph.

EXAMPLE 3.6. The matrix

$$A = \begin{bmatrix} 5 & 1 & 3 \\ 3 & 2 & 2 \\ 0 & 7 & 1 \end{bmatrix}$$

has no nontrivial factorizations $A = RS$ into nonnegative integer matrices R, S of size 3×3 or smaller, and so is not strongly shift equivalent in one step to any other 3×3 or smaller matrix, other than versions of itself with permuted indices.

However, A has characteristic polynomial $\lambda^2(\lambda - 8)$, which up to factors of λ is the characteristic polynomial of the full 8-shift (8, as a 1×1 matrix), namely $\lambda - 8$. It can be shown that a square integer matrix with the characteristic polynomial of a full k-shift (up to factors of λ) is actually shift equivalent to the k-shift. In the present case, $A^2 = RS$, $SR = 8^2$, $SA = 8S$, $AR = R8$ for $R = [4\ \ 3\ \ 3]^t$, $S = [7\ \ 7\ \ 5]$. By Williams's conjecture, A should be strongly shift equivalent to 8. Is it?

Theorem 3.2 provides an answer: Observe that the first row of A dominates the second row except in the $(1,2)$ position. Thus it is worthwhile examining $B = H^{-1}AH$ for $H = U_3[1,2]$, and indeed,

$$B = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 2 & 2 \\ 0 & 7 & 1 \end{bmatrix},$$

which is nonnegative with $b_{12} > 0$. Therefore $A \to B$.

It is now easy to continue: Observe that the second row of B dominates the first, so that $B \to C = G^{-1}BG$ for

$$G = U_3[2,1] \quad \text{and} \quad C = \begin{bmatrix} 3 & 1 & 1 \\ 5 & 4 & 1 \\ 7 & 7 & 1 \end{bmatrix}.$$

In fact $B \approx_1 C$. Thus A and C are in the same component of SDG_A, and by Theorem 3.2, $A \approx C$.

Now,

$$C = \begin{bmatrix} 3 & 1 & 1 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

These factors multiplied in the other order give

$$\begin{bmatrix} 5 & 3 \\ 5 & 3 \end{bmatrix},$$

which also factors as

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} [5 \quad 3].$$

These factors multipled in the other order give the 1×1 matrix 8.

Therefore $A \approx 8$, a conclusion consistent with Williams's conjecture.

The full shear digraph of A has many thousands of vertices, and in fact the component of A has 6410 vertices.

4. SHEAR DIGRAPHS IN DIMENSION 2

LEMMA 4.1. *In the shear digraph of an $n \times n$ nonnegative integer matrix, each vertex has indegree and outdegree at most $n(n-1)/2$ (which is 1, in the case $n = 2$).*

Proof. For the case $n = 2$, suppose $A \to B$ by $B = U^{-1}AU$, where $U = U_2[i, j]$. Observe that b_{ij} is the ith row sum of A minus the jth row sum. Further, by the definition of $A \to B$, $b_{ij} > 0$. Therefore the cases $i = 1$, $j = 2$ and $i = 2$, $j = 1$ are not simultaneously possible. For general n and given A, $U_n[k, l]$ and $U_n[l, k]$ cannot both yield nonnegative B with $A \to B$, by the same reasoning applied to the 2×2 principal submatrix for rows and columns k, l. Therefore a bound on the outdegree of A is the number of pairs $\{k, l\}$, namely $n(n-1)/2$. A transposed argument applies to the indegree. ∎

(The bound can easily be attained, for example if each row of A dominates the next entrywise.)

COROLLARY 4.2. *The shear digraph of a 2×2 nonnegative integer matrix is a disjoint union of cycles and linear digraphs.*

Here a *linear digraph* is a finite or infinite directed graph isomorphic to an interval of the integers with edges corresponding to the successor relation.

LEMMA 4.3. *For monic $p(\lambda) \in \mathbf{Z}[\lambda]$ of degree 2, if some component of $\mathrm{SDG}_{p(\lambda)}$ is a cycle, then either*

(a) *$p(\lambda)$ is irreducible in $\mathbf{Z}[\lambda]$ and the cycle is of length > 1, or else*

(b) $p(\lambda)$ *is a square and the cycle is of length* 1 (*a loop*), *with the vertex having the form*

$$\begin{bmatrix} a & b \\ 0 & a \end{bmatrix} \quad \textit{or the transpose.}$$

Proof. Suppose A is a vertex in an n-cycle. The successive vertices around the cycle are obtained by successive conjugations of A by unit shears $H_1,\dots,H_n$. Let $P = H_1 \cdots H_n$. Then P is a nonnegative integer matrix invertible over the integers, and $A = P^{-1}AP$. Since A and P commute, A and P are simultaneously reducible to Jordan form. In fact, since A and P are 2×2 and evidently neither is scalar, each is a linear combination of the other and I over the rationals, and their characteristic polynomials are either both reducible or both irreducible in $\mathbf{Z}[\lambda]$.

Suppose both characteristic polynomials are reducible. Then the eigenvalues of P are invertible rational integers, i.e., ± 1. If 1 and -1 were both eigenvalues, then P would be diagonalizable and $P^2 = I$—not the case here, since $P \geqslant$ each H_i entrywise. Since $P \geqslant 0$, both eigenvalues must then be 1; since P is not scalar, P must have nondiagonal Jordan form and satisfy $(P-I)^2 = 0$. Since $P \geqslant I$, P must actually be triangular. Since $P \geqslant H_i$ for all i, the H_i must be all the same unit shear, with P being a power of it, and the H_i commute with A. Therefore the cycle is a 1-cycle and A and $p(\lambda)$ have the stated forms.

Suppose both characteristic polynomials are irreducible. Then the cycle length cannot be 1, as then A would commute with a unit shear and so would be triangular, hence reducible. ■

The following concept will be convenient:

Definition 4.4 Let us call a pair $\{r, s\}$ of real numbers *almost positive* if at least one of r, s is nonnegative and $r + s + rs \geqslant 0$.

Thus we may speak of a polynomial of degree 2 with almost positive roots, or a 2×2 matrix with almost positive spectrum.

The pairs $(r, s) \in \mathbf{R}^2$ with $\{r, s\}$ almost positive form a convex subset of the half plane $x + y \geqslant 0$ containing the first quadrant.

Observation 4.5. *For a* 2×2 *real matrix* A *with characteristic polynomial* $p(\lambda)$ *and real eigenvalues, the following conditions are equivalent*:

(i) *the spectrum of* A *is almost positive*;
(ii) $\operatorname{trace} A \geqslant 0$ *and* $\det A + \operatorname{trace} A \geqslant 0$;

(iii) $\det(A+I)\geqslant 1$ *and* A *has at least one nonnegative eigenvalue*;
(iv) $p'(0)\leqslant 0$ *and* $p(-1)\geqslant 1$.

Condition (ii) will be the one of direct interest; by its first part, its second part is best viewed as the requirement that $\det A$ be at least as large as $-\operatorname{trace} A$. Conditions (iii) and (iv) can be regarded as technical restatements of this property.

Observe that a nonnegative 2×2 integer matrix with nonnegative determinant (the case treated in [2]) satisfies these equivalent conditions. (Recall that a 2×2 nonnegative matrix has real eigenvalues.)

THEOREM 4.6. *Let* $p(\lambda)\in \mathbf{Z}[\lambda]$ *be monic of degree* 2 *with* $\mathrm{SDG}_{p(\lambda)}$ *nonempty. Then these conditions are equivalent*:

(a) *every component of* $\mathrm{SDG}_{p(\lambda)}$ *is a cycle of length two or greater*;
(b) $p(\lambda)$ *is irreducible with almost positive roots.*

Proof. For (a) $\Rightarrow$ (b): $p(\lambda)$ is irreducible by Lemma 4.3. To show that $p(-1)\geqslant 1$, write $p(\lambda)=\lambda^2-t\lambda+d$. Since $\mathrm{SDG}_{p(\lambda)}$ is nonempty, the roots are real and $t\geqslant 0$. If $t+d<0$, then the companion matrix

$$\begin{bmatrix} 0 & -d \\ 1 & t \end{bmatrix}$$

has outdegree 0, in contradiction of (a).

For (b) $\Rightarrow$ (a): By Proposition 2.5 the shear digraph is finite; therefore it suffices to show that the outdegree of every vertex is 1. Since $p(\lambda)$ is irreducible, no matrix with characteristic polynomial $p(\lambda)$ is triangular. Let A be a vertex and let $B=U^{-1}AU$ for $U=U_2[i,j]$. Then as noted above, $b_{ij}=(a_{i1}+a_{i2})-(a_{j1}+a_{j2})$, a difference of row sums. The row sums cannot be equal, as in that case the 2×2 matrix B would be triangular. Without loss of generality, suppose that row 1 of A has the greater row sum and row 2 the lesser and set $i=1$, $j=2$. To show $A\to B$, since $a_{12}>0$ and $b_{12}>0$, it suffices to show $B\geqslant 0$ by noting that $b_{21}=a_{21}>0$ and appealing to the following fact:

LEMMA 4.7. *Let* B *be a* 2×2 *integer matrix with nonnegative trace, nonnegative off-diagonal elements, and characteristic polynomial* $p(\lambda)$ *with* $p(-1)\geqslant 1$. *Then* $B\geqslant 0$.

Proof. The discriminant of B is $(b_{11}-b_{22})^2+4b_{12}b_{21}\geqslant 0$; therefore B has real eigenvalues. By Observation 4.5, $1\leqslant \det(B+I)=(b_{11}+1)(b_{22}+1)-b_{12}b_{21}\leqslant(b_{11}+1)(b_{22}+1)$. Then either $b_{11}, b_{22}\geqslant 0$ as desired, or else $b_{11}, b_{22}\leqslant -2$, an impossibility if the trace is to be nonnegative. ∎

Example 4.8 (A shear digraph with almost positive eigenvalues).

$$\begin{bmatrix}11&2\\3&5\end{bmatrix}\to\begin{bmatrix}8&5\\3&8\end{bmatrix}\to\begin{bmatrix}5&2\\3&11\end{bmatrix}\to\begin{bmatrix}7&2\\7&9\end{bmatrix}\to\begin{bmatrix}9&2\\7&7\end{bmatrix}\to\begin{bmatrix}11&2\\3&5\end{bmatrix}$$

$$\begin{bmatrix}11&6\\1&5\end{bmatrix}\to\begin{bmatrix}10&11\\1&6\end{bmatrix}\to\begin{bmatrix}9&14\\1&7\end{bmatrix}\to\begin{bmatrix}8&15\\1&8\end{bmatrix}\to\begin{bmatrix}7&14\\1&9\end{bmatrix}\to\begin{bmatrix}6&11\\1&10\end{bmatrix}$$
$$\to\begin{bmatrix}5&6\\1&11\end{bmatrix}\to\begin{bmatrix}11&6\\1&5\end{bmatrix}$$

$$\begin{bmatrix}5&3\\2&11\end{bmatrix}\to\begin{bmatrix}8&3\\5&8\end{bmatrix}\to\begin{bmatrix}11&3\\2&5\end{bmatrix}\to\begin{bmatrix}9&7\\2&7\end{bmatrix}\to\begin{bmatrix}7&7\\2&9\end{bmatrix}\to\begin{bmatrix}5&3\\2&11\end{bmatrix}$$

$$\begin{bmatrix}5&1\\6&11\end{bmatrix}\to\begin{bmatrix}6&1\\11&10\end{bmatrix}\to\begin{bmatrix}7&1\\14&9\end{bmatrix}\to\begin{bmatrix}8&1\\15&8\end{bmatrix}\to\begin{bmatrix}9&1\\14&7\end{bmatrix}\to\begin{bmatrix}10&1\\11&6\end{bmatrix}$$
$$\to\begin{bmatrix}11&1\\6&5\end{bmatrix}\to\begin{bmatrix}5&1\\6&11\end{bmatrix}.$$

Remark 4.9.

(1) Starting from a 2×2 nonnegative integer matrix A meeting the conditions of Proposition 4.6, it is a simple matter to use the procedure of the proof repeatedly to find the cycle of A in SDG_A.

(2) The cycles of A^t, $P^{-1}AP$, and $P^{-1}A^tP$ may or may not coincide with the cycle of A.

(3) If T is a triangular 2×2 nonnegative integer matrix, then its shear digraph can be infinite, for example,

$$T=\begin{bmatrix}2&1\\0&1\end{bmatrix}.$$

(4) If $p(\lambda)$ is reducible as a square, then as in Lemma 4.3, $\mathrm{SDG}_{p(\lambda)}$ consists of infinitely many components, each a loop (cycle of length 1). If $p(\lambda)$ is reducible but not a square, then $\mathrm{SDG}_{p(\lambda)}$ consists of finitely many finite linear components and, if both eigenvalues are positive, finitely many

components each a one-sided-infinite linear digraph consisting of triangular matrices.

(5) If $p(\lambda)$ is irreducible but $p(-1)<1$, then in general $\mathrm{SDG}_{p(\lambda)}$ consists of finitely many components each of which is either a cycle of length at least two or a linear digraph (possibly of one vertex).

(6) The existence of cycles corresponds to periodicity in the continued-fraction expansion of eigenvalues (see [2]), which holds for eigenvalues of degree 2 only. The decomposition into cycles does not in general extend to shear digraphs of 3×3 matrices or larger.

5. SPECTRAL TRANSLATION

Let us call the process of replacing a square matrix A by $A+kI$ for some $k\in\mathbf{Z}$ *spectral translation*, as the eigenvalues of A are thereby shifted by k.

At first, this process would seem to have no regular properties with respect to strong shift equivalence:

EXAMPLE 5.1. Let

$$A=\begin{bmatrix}4 & 5\\ 3 & 4\end{bmatrix},\qquad B=\begin{bmatrix}1 & 2\\ 3 & 7\end{bmatrix},\qquad C=\begin{bmatrix}4 & 15\\ 1 & 4\end{bmatrix}.$$

Then

(i) $A\approx_1 B$,

(ii) $A+10I\approx B+10I$, but only in many steps with 3×3 intermediate matrices,

(iii) $A+I\approx_1 C+I$,

(iv) A and C themselves are not even shift equivalent, nor are $A+10I$ and $C+10I$.

[For (iv), observe that $A^n=RS$ implies that $\det S=\pm 1$, while the matrix equation $SA=CS$ leads to a system of linear equations that show $\det S\equiv\pm 2$ or 0 (mod 5); the case of $A+10I$ and $C+10I$ is similar.]

However, one fact is evident:

OBSERVATION 5.2. *If A, B are nonnegative and $k\geqslant 0$, then $A\to B$ if and only if $(A+kI)\to(B+kI)$.*

Indeed, similarity via $U_n[i,j]$ is preserved under spectral translation, as are the off-diagonal elements needed for condition (ii) *of Definition* 2.3.

Remark 5.3. If trace $A \geqslant 0$, then spectral translation of A by a positive k increases both the trace and the determinant.

Lemma 5.4. *If A is a 2×2 nonnegative integer matrix with almost positive spectrum, and if $k \geqslant 0$ in $\mathbf{Z}$, then spectral translation by k is a graph isomorphism on SDG_A to SDG_{A+kI}.*

Proof. For a square matrix M, write $\phi_k(M) = M + kI$. By Observation 5.2, it suffices to show that ϕ_k and ϕ_{-k} give a one-to-one correspondence between SDG_A and SDG_{A+kI}. The only question is whether $\phi_{-k}(B)$ must be nonnegative for $B \in \mathrm{SDG}_{A+kI}$. But $\phi_{-k}(B)$ has the characteristic polynomial of A and nonnegative off-diagonal entries, so that its nonnegativity follows from Observation 4.5 and Lemma 4.7. ∎

Lemma 5.5. *If A is a nontriangular 2×2 nonnegative integer matrix with* $\det A \geqslant -\operatorname{trace} A$, *then the component of A in its shear digraph consists of all nonnegative matrices similar to A by an integer matrix of determinant* 1.

Proof. For $\det A \geqslant 1$, this is the import of Lemma 2.1 of [2]. (Note that A has positive entries.) If $\det A < 1$, then we may spectrally translate A by a sufficiently large positive integer k that a matrix of determinant $\geqslant 1$ is obtained, apply the earlier case, translate back, and use Lemma 5.4. ∎

Theorem 5.6. *If A and B are 2×2 nontriangular nonnegative integer matrices similar over the integers and* $\det A + \operatorname{trace} A \geqslant 0$, *then A and B are strongly shift equivalent.*

Proof. If A and B are similar by an integer matrix of determinant 1, then Lemma 5.5 and hence Corollary 4.2 apply. If A and B are similar by an integer matrix of determinant -1, then A is similar by an integer matrix of determinant 1 to $H^{-1}BH$, where H is the permutation matrix of a transposition, and we have $A \approx H^{-1}BH \approx B$. ∎

Discussion 5.7. We now have this picture: If A is a 2×2 nonnegative integer matrix with $\det A \geqslant -\operatorname{trace} A$ and irreducible characteristic polynomial, then

(a) the SL(2, Z)-similarity classes of nonnegative matrices with the characteristic polynomial of A are simply the components of SDG_A, all of which are cycles;

(b) each GL(2, Z)-similarity class consists of either one or two cycles, depending on whether similarity by an integer matrix of determinant -1 preserves the class or not;

(c) each strong-shift-equivalence class consists of one or more GL(2, Z)-similarity classes (each corresponding to an ideal class of $\mathbf{Z}[A]$ [23]);

(d) each shift-equivalence class consists of one or more strong-shift-equivalence classes (according to Williams's conjecture, just one.)

In particular, classes of these various kinds are unions of cycles.

Especially for nonnegative 2×2 integer matrices A such that $\det A < -\operatorname{trace} A$, the following concept is helpful:

DEFINITION 5.8. The *eventual shear digraph* $\mathrm{ESDG}_{p(\lambda)}$ of a polynomial $p(\lambda)\in \mathbf{Z}[\lambda]$, monic of degree 2 with real roots, is (the isomorphism type of) the shear digraph of $p(\lambda-k)$ for sufficiently high $k\geqslant 0$, $k\in\mathbf{Z}$. For a 2×2 nonnegative integer matrix A, its eventual shear digraph ESDG_A is that of its characteristic polynomial, or in other words, (the isomorphism type of) SDG_{A+kI} for sufficiently high $k\geqslant 0$.

Indeed, $q(\lambda)=p(\lambda-k)$ has almost positive spectrum for sufficiently high k.

THEOREM 5.9. *Let A be a 2×2 nonnegative integer matrix with irrational eigenvalues. Then the eventual shear digraph ESDG_A of A is a disjoint union of cycles. If A is almost positive, SDG_A is isomorphic to ESDG_A; otherwise SDG_A is a proper subgraph of ESDG_A.*

Proof. Choose any $k\geqslant 0$ such that $\det(A+kI)+\operatorname{trace}(A+kI)\geqslant 0$. Then $\mathrm{ESDG}_A\cong\mathrm{SDG}_{A+kI}$. By Proposition 4.6, SDG_{A+kI} is a disjoint union of cycles. If A is not almost positive, when the vertices of SDG_{A+kI} are spectrally translated back by $-k$, some of the resulting matrices will have negative diagonal entries; in fact,

$$\begin{bmatrix} -1 & -(d+t+1) \\ 1 & t+1 \end{bmatrix}$$

is one such, where $d=\det A$ and $t=\operatorname{trace}A$. SDG_A consists of the remaining matrices, connected as in SDG_{A+kI}. ■

In effect, some vertices of a cycle in SDG_{A+kI} may become "submerged" in SDG_A, and the remaining "islands" are linear digraphs. According to

Williams's conjecture, matrices in different islands from the same cycle should still be strongly shift equivalent, even though the method of Section 3 no longer is applicable. And indeed, computer experiments in cases examined successfully confirm the strong shift equivalence but suggest that 4×4 intermediate matrices can be necessary.

EXAMPLE 5.10. $\mathrm{SDG}_{p(\lambda)}$ for $p(\lambda)=\lambda^2-4\lambda-17$. Half the graph is shown below; the remaining vertices are transposes. Commas separate components that are islands from the same component of the eventual shear digraph, which is obtainable using $k\geqslant 2$ and consists of two 14-cycles and two 4-cycles:

$$\begin{bmatrix}3&4\\5&1\end{bmatrix},\ \begin{bmatrix}0&1\\17&4\end{bmatrix}\to\begin{bmatrix}1&1\\20&3\end{bmatrix}\to\begin{bmatrix}2&1\\21&2\end{bmatrix}\to\begin{bmatrix}3&1\\20&1\end{bmatrix}\to\begin{bmatrix}4&1\\17&0\end{bmatrix},$$

$$\begin{bmatrix}1&4\\5&3\end{bmatrix},\ \begin{bmatrix}2&7\\3&2\end{bmatrix};\ \begin{bmatrix}1&2\\10&3\end{bmatrix}\to\begin{bmatrix}3&2\\10&1\end{bmatrix}$$

REFERENCES

1 R. L. Adler, D. Coppersmith, and M. Hassner, Algorithms for sliding block codes, *IEEE Trans. Inform. Theory* IT-29:5–22 (1983).

2 K. A. Baker, Strong shift equivalence of 2 by 2 matrices of nonnegative integers, *Ergodic Theory Dynamical Systems* 3:501–508 (1983).

3 J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, American Elsevier, New York, 1976.

4 R. Bowen, Markov partitions for axiom-A diffeomorphisms, *Amer. J. Math.* 92:725–747 (1970).

5 R. Bowen, Markov partitions and minimal sets for axiom A diffeomorphisms, *Amer. J. Math.* 92:907–918 (1970).

6 R. Bowen and O. Lanford, Zeta functions of restrictions of the shift transformation, in *Proceedings of Symposia in Pure Mathematics*, Vol. 14, Amer. Math. Soc., Providence, R.I., 1970, pp. 43–50.

7 M. Boyle, Shift equivalence and the Jordan form away from zero, *Ergodic Theory Dynamical Systems* 4:367–380 (1984).

8 J. Cuntz and W. Krieger, Topological Markov chains with dicyclic dimension groups, *J. Reine Angew. Math.* 320:44–51 (1980).

9 M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic Theory on Compact Spaces*, Lecture Notes in Mathematics No. 527, Springer, New York, 1976.

10 J. Franks, *Homology and Dynamical Systems*, Amer. Math. Soc. CBMS Conf. Ser. 49, 1982.

11 F. R. Gantmacher, *The Theory of Matrices*, Vol. II, Chelsea, New York, 1960.

12 M. Hassner, A Nonprobabilistic Source and Channel Coding Theory, Ph.D. Thesis, Univ. of California, Los Angeles, 1980.
13 K. H. Kim and F. W. Roush, Some results on decidability of shift equivalence, *J. Combin. Inform. System Sci.* 4:123–146 (1979).
14 K. H. Kim and F. W. Roush, Decidability of shift equivalence, to appear.
15 W. Krieger, On dimension functions and topological Markov chains, *Invent. Math.* 63:239–250 (1980).
16 W. Krieger, On certain notions of equivalence for topological Markov chains, to appear.
17 D. A. Lind, Entropies and factorizations of topological Markov shifts, *Bull. Amer. Math. Soc. (N.S.)* 9:219–222 (1983).
18 D. A. Lind, The entropies of topological Markov shifts and a related class of algebraic integers, *Ergodic Theory Dynamical Systems* 4:283–300 (1984).
19 M. Morse and G. Hedlund, Symbolic dynamics, *Amer. J. Math.* 60:815–866 (1938).
20 M. Newman, *Integral Matrices*, Academic, New York, 1972.
21 O. T. O'Meara, The automorphisms of the linear groups over any integral domain, *J. Reine Angew. Math.* 223:55–100 (1966).
22 W. Parry, Intrinsic Markov chains, *Trans. Amer. Math. Soc.* 112:55–66 (1964).
23 O. Taussky, On a theorem of Latimer and McDuffee, *Canad. J. Math.* 1:300–302 (1949).
24 R. F. Williams, Classification of subshifts of finite type, *Ann. of Math.* 98:120–153 (1973); Errata, *ibid.* 99:380–381 (1974).

Received 2 June 1986; revised 31 July 1986